Option Hedging with Risk Averse Reinforcement Learning

Edoardo Vittori

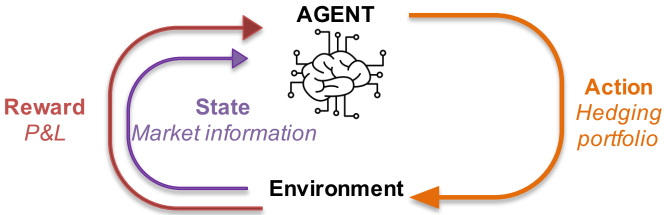
Based on a work done with Michele Trapletti and Marcello Restelli

Outline

- Option Hedging
- Reinforcement Learning Intro
- State of the Art
- Risk Averse RL
- Experimental results
- Conclusions

Option hedging: trading the *underlying asset* in order to minimize the price swings generated by the option (controlling risk).

Reinforcement Learning Intro



- the action $a_t \in [0,1]$ the hedging portfolio
- the state $s_t = (S_t, C_t, \frac{\partial C_t}{\partial S_t}, a_{t-1})$
- the reward $R(s_t, a_t) = C_{t+1}(S_{t+1}) - C_t(S_t) - a_t \cdot (S_{t+1} - S_t) - c(n)$
- transaction costs $c(n) = 0.05 \cdot (|n| + 0.01n^2)$, $n = a_t - a_{t-1}$

Reinforcement Learning Intro

- Returns

$$G(\tau) = \sum_{t=0}^{\infty} \gamma^t R_t$$

- Action-Value function

$$Q_\pi(s, a) = \mathop{\mathbb{E}}_{\tau\sim\pi}[G(\tau)|s_0 = s, a_0 = a]$$

- Objective

$$J = \max_{\pi} \mathop{\mathbb{E}}_{\tau\sim\pi}[G(\tau))]$$

- Policy Search vs Value Based approaches

State of the Art

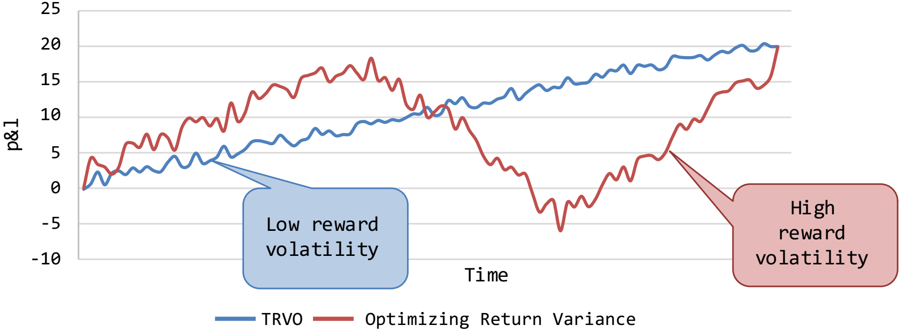
Reinforcement Learning in Finance

- RL in Hedging
  - (Halperin, 2017)
  - (Halperin, 2019)
  - (Kolm and Ritter, 2019a)
  - (Kolm and Ritter, 2019b)
  - (Buehler et al., 2019)
  - (Cao et al., 2019)

Risk Averse Reinforcement Learning

- Reward volatility
  - (Bisi et al., 2020)
- Utility based
  - (Moldovan and Abbeel, 2012)
  - (Shen et al., 2014)
- Coherent Risk Measures
  - (Morimura et al., 2010)
  - (Tamar et al., 2017)
  - (Chow et al., 2017)
- Variance of the returns
  - (Sobel, 1982)
  - (Tamar and Mannor, 2013)
  - (Prashanth and Ghavamzadeh, 2014)

Variance/Volatility relation



$$\sigma_\pi^2 \leqslant \frac{\nu_\pi^2}{(1-\gamma)^2}$$

Reward Volatility

- Reward volatility

$$\nu_\pi^2 = (1-\gamma) \mathop{\mathbb{E}}_{\substack{s_0 \sim \mu \\ a_t \sim \pi(\cdot|s_t) \\ s_{t+1} \sim \mathcal{P}(\cdot|s_t, a_t)}} \left[\sum_{t=0}^{\infty} \gamma^t \left(\mathcal{R}(s_t, a_t) - J_\pi \right)^2 \right]$$

- Mean-volatility objective

$$\eta_\pi := J_\pi - \lambda \nu_\pi^2$$

- Trust Region Volatility Optimization-TRVO (Bisi et al., 2020)

Financial environment

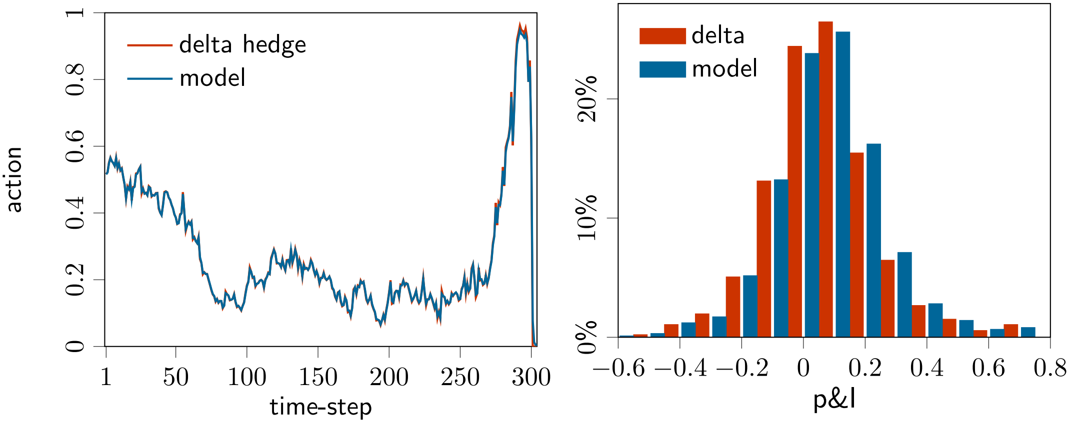
Vanilla call option

- time to maturity = 60 days
- unitary notional
- implied volatility = 20%
- interest rates = 0
- $K (= S_0) = 100$
- starting price (ATM) option ~ 3.24
- starting delta $= 0.5$

Simulated Market

- geometric brownian motion $dS_t = \mu S_t dt + \sigma S_t dW_t$
- no drift
- $\sigma = 20\%$
- $S_0 = 100$
- 5 time steps per day

Experimental Results: no Transaction Costs

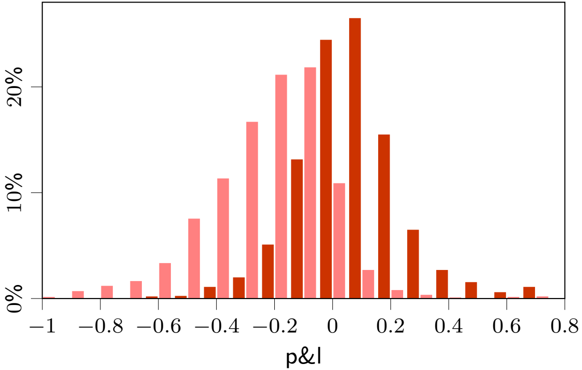


⇒ delta hedge with no costs → average p&l $\sim$ 0, volatility $\sim$ 0.16

Adding costs

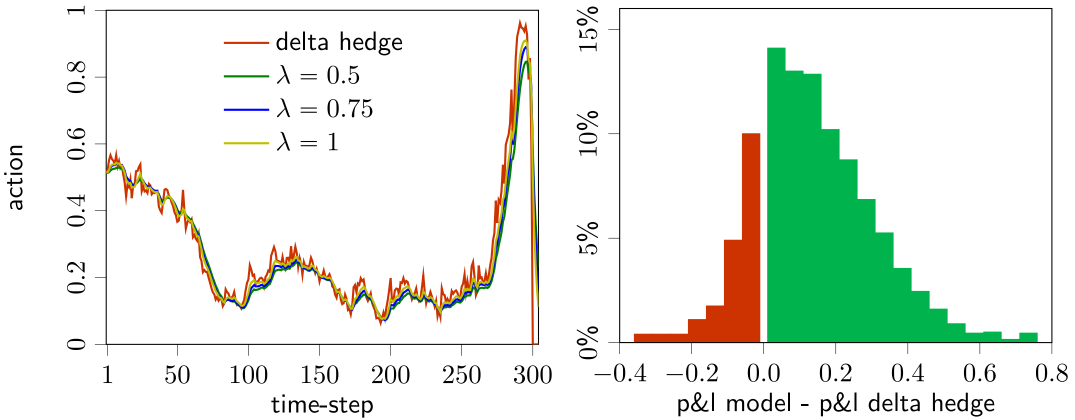
We considered h.c.$\sim 0.05|a|$, $\sim$ the Euro Stoxx 50 or FTSE MIB future.

- More liquid listed products (S&P 500) have lower minimal costs
- Less liquid listed or OTC (vanilla, flow) instruments have significantly higher costs



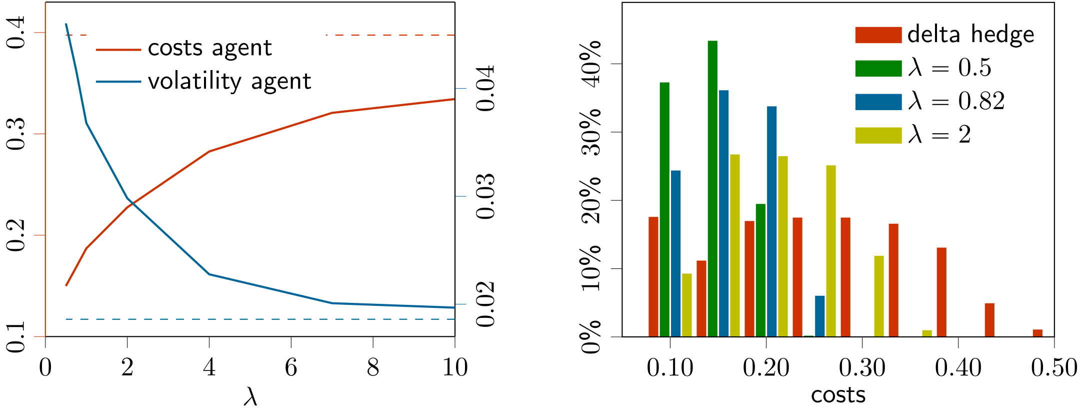
- The agent now has something to optimize: costs vs. volatility
- Costs give a role to the risk aversion factor

Experimental Results: with Transaction Costs

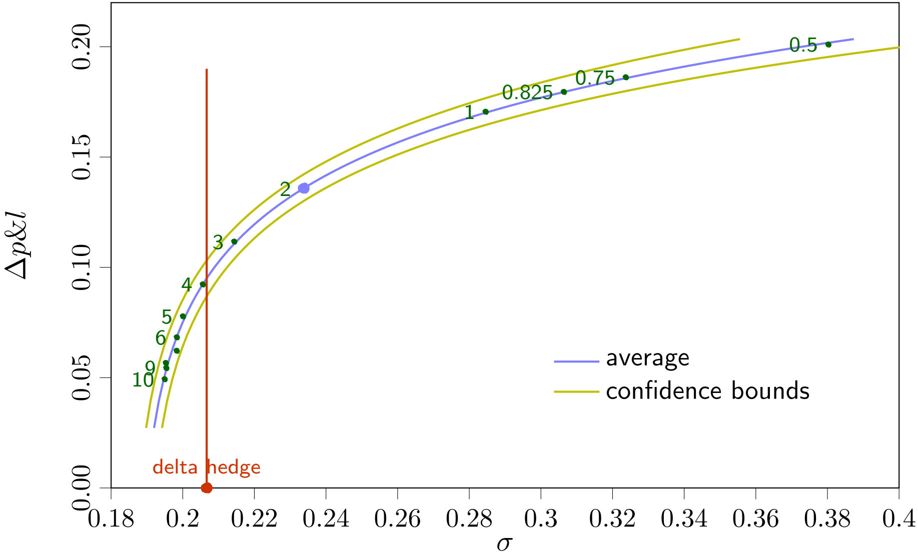


⇒ delta hedge with no costs → average p&l ∼ -0.3, volatility ∼ 0.18

Experimental Results: Cost Analysis



Experimental Results: Pareto Plot



Conclusions

Contributions

- Proved experimentally that the hedging strategy learnt by the model dominates the delta hedge

Future works

- Extend to more complex derivatives
- Extend to a portfolio of options
- Decide not only how much but also when to hedge

Contacts

- Edoardo Vittori - edoardo.vittori@polimi.it
- Michele Trapletti - michele.trapletti@intesasanpaolo.com
- Marcello Restelli - marcello.restelli@polimi.it

Thank You for Your Attention!

References

Lorenzo Bisi, Luca Sabbioni, Edoardo Vittori, Matteo Papini, and Marcello Restelli. Risk-averse trust region optimization for reward-volatility reduction. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4583–4589. International Joint Conferences on Artificial Intelligence Organization, 7 2020. Special Track on AI in FinTech.

Hans Buehler, Lukas Gonon, Josef Teichmann, and Ben Wood. Deep hedging. *Quantitative Finance*, pages 1–21, 2019.

Jay Cao, Jacky Chen, John C. Hull, and Zissis Poulos. Deep hedging of derivatives using reinforcement learning. *Available at SSRN*, 2019.

Yinlam Chow, Mohammad Ghavamzadeh, Lucas Janson, and Marco Pavone. Risk-constrained reinforcement learning with percentile risk criteria. *JMLR*, 18(1):6070–6120, 2017.

Igor Halperin. Qlbs: Q-learner in the black-scholes (-merton) worlds. *The Journal of Derivatives*, 2017.

Igor Halperin. The qlbs q-learner goes nuqlear: fitted q iteration, inverse rl, and option portfolios. *Quantitative Finance*, pages 1–11, 2019.

Petter N Kolm and Gordon Ritter. Dynamic replication and hedging: A reinforcement learning approach. *The Journal of Financial Data Science*, 1(1):159–171, 2019a.

Petter N Kolm and Gordon Ritter. Modern perspectives on reinforcement learning in finance. *Modern Perspectives on Reinforcement Learning in Finance (September 6, 2019). The Journal of Machine Learning in Finance*, 1(1), 2019b.

References (cont.)

Teodor M. Moldovan and Pieter Abbeel. Risk aversion in Markov decision processes via near optimal Chernoff bounds. In *NeurIPS*, pages 3131–3139, 2012.

Tetsuro Morimura, Masashi Sugiyama, Hisashi Kashima, Hirotaka Hachiya, and Toshiyuki Tanaka. Nonparametric return distribution approximation for reinforcement learning. In *ICML*, 2010.

L. A. Prashanth and Mohammad Ghavamzadeh. Actor-critic algorithms for risk-sensitive reinforcement learning. *arXiv preprint arXiv:1403.6530*, 2014.

Yun Shen, Ruihong Huang, Chang Yan, and Klaus Obermayer. Risk-averse reinforcement learning for algorithmic trading. pages 391–398, March 2014. doi: 10.1109/CIFEr.2014.6924100.

Matthew J. Sobel. The variance of discounted Markov decision processes. *Journal of Applied Probability*, 19(4): 794–802, 1982.

Aviv Tamar and Shie Mannor. Variance adjusted actor critic algorithms. *arXiv preprint arXiv:1310.3697*, 2013.

Aviv Tamar, Yinlam Chow, Mohammad Ghavamzadeh, and Shie Mannor. Sequential Decision Making With Coherent Risk. *IEEE Transactions on Automatic Control*, 62(7):3323–3338, July 2017. ISSN 0018-9286, 1558-2523. doi: 10.1109/TAC.2016.2644871. URL `http://ieeexplore.ieee.org/document/7797146/`.